The accuracy of indirect measurements and model calibration by falsification

J Wendelstorf

Cite as: preprint (Hochschulschrift) http://doi.org/10.21268/20190515-4

Priority date: April, 3rd 2019

Clausthal University of Technology
Robert-Koch-Str. 42
D-38678 Clausthal-Zellerfeld
Germany

E-mail: jens.wendelstorf@tu-clausthal.de

Abstract. Scientific measurements should deliver quantitative values and error specifications. This requirement is also valid for indirect measurements, were model parameters are deduced from direct measurements. The regression (fitting) of parameters should include a specification of their accuracies. This is difficult for the case of two or more parameters. An individual direct measurement can be quantified by a numerical interval. It represents all available information in the frequent case of a predominating systematic error. For models with more than one parameter it is not constructive to quantify the parameters by intervals as these enclose parameter sets not supported by the measurements. One has to determine the region of possible parameters by a special kind of membership function. The model falsification indicator function (MFIF) F determines the region of possible parameters not falsified by a given set of measurements. Alternatively it provides a measure for the fraction of outliers in the experimental data set. The MFIF can be used for model identification and calibration (regression analysis). Being a positive realization of Popper's falsification approach, it is useful for the determination of structural and practical parameter identifiability, model sloppiness, outlier detection and the discussion of model validation questions. The MFIF is applicable to any model and scientific discipline. It is exemplified and visualized for the simplest possible (linear) case and an easily reproducible nonlinear example. Being simple to implement, this new mathematical tool demands the attention of all scientists.

Keywords:
accuracy, falsification, indirect measurement, parameter, model identification

1. Introduction

For a complex system being the subject of a scientific study, the direct measurement of all physical parameters is often not possible. Many processes only permit the direct measurement of quantities not identical with the main parameters of the system. The scientific method then demands for a mathematical description of the process, a model Y. While the systematic and random errors of direct measurements can be quantified easily [1], this is not the case for models with more than one parameter. For the frequent case of predominating systematic errors in the direct measurements, statistics are neither required nor helpful. To understand these incredible facts, some mathematical formalism is required.

A model Y represents a process controlled by n_i measurable quantities, the inputs x_i, and allows the computation of n_j measurable quantities, the outputs y_j. Such models may incorporate n_k parameters $\boldsymbol{p} = (p_1, \ldots, p_{n_\mathrm{k}})$ which are not directly measurable by another process. The determination (indirect measurement) of these parameters using measurements with correctly specified accuracy is the objective of this study.

The solution delivers a mathematical formulation of Karl Popper's scientific epistemology ([2], for a more philosophical discussion see e.g. [3]). Summarily it is a positive implementation of Richard Feynman's statement from 1964: *It doesn't matter how beautiful your theory is, it doesn't matter how smart you are. If it doesn't agree with experiment, it is wrong.*

There are standard methods for the determination of a parameter set representing the *best fit* of the data to the model [4, 5, 6]. Based on statistical assumptions on the precision of the measurements, there is a variety of goodness of fit functions $G(\boldsymbol{p})$ to be minimized by (global) optimization algorithms. This is known and still under refinement as the method of least squares [7, 8]. For a given set of measurements, it allows to calculate a parameter vector

$$\boldsymbol{p}_0 = \arg\min G(\boldsymbol{p}) \tag{1}$$

providing the *best fit* of the measurements to the model. Only for the case of predominating random errors in the underlying direct measurements, $\boldsymbol{p}_0$ is superior to other valid parameter vectors [1]. Additionally, under special conditions, it allows to calculate a confidence region in the parameter space,

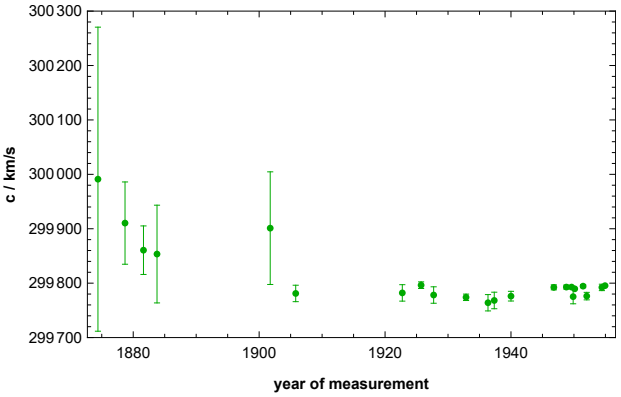

Figure 1. History of the measurements of the velocity of light c (data from [16]).

where the measurements are in reasonable agreement with the predictions of the model. The numerical approximation of these confidence regions is reviewed in [9, 10, 11]. The results depend on the statistical assumptions about the precision of the underlying measurements.

As an alternative, model falsification is also applicable to real world problems, e.g. the error domain model falsification (EDMF) used in civil engineering [12]. Interval analysis [13] provides the mathematical foundation for methods free of statistical assumptions, only the specification of bounds for the measured values is required [14, 15]. For lowering its entry threshold, this bounded error approach may benefit from an easy to implement basis, as aspired here.

The fundamental question is: How can we obtain the region in the parameter space of the model, where the model is not falsified by the measurements? Is this a quest, *too difficult to attain* [3], or is it computationally challenging only? There may be measurements not in agreement with the range of values predicted by the model using the *best fit* model parameter set $\boldsymbol{p}_0$. These measurements will not falsify the model, if there is at least one other parameter set for which the model's predictions are in accordance with all measurements (within their accuracy and precision limits). Alternatively, if we can not trust in all of the measurements, under the assumption of a correct model and a correct parameter set, we have to determine the outliers.

Another question is the justification of statistical assumptions in the case of relevant measurement

accuracies (systematic error). From an *a posteriori* point of view (see figure 1), any initial assumption on the probability distribution within the bounds given by the measurement accuracies is challenged later by more accurate measurements. Regarding systematic error, the true probability distribution function (PDF) has to be peaked at the true value, which is unknown in principle. As can be observed in figure 1, there is also a slight error underestimation tendency even for reliable measurements undertaken by diligent scientists. Approximating the PDF of a systematic error is like approximating a Dirac δ-function for an unknown position of its origin.

In the following section, an example - as simple as possible - motivates a novel approach, as simple as (currently) possible. In section 3, the general definition of the *model falsification indicator function* (MFIF) F is discussed and in section 4, a non-linear example demonstrates the general applicability of the MFIF approach. As motivated in section 5, the application is possible within the framework of existing algorithms, computational tools and resources by direct visualization of F or using it for sampling the region of possible model parameters $\mathbf{R}$.

2. An example, as simple as possible

Regarding model falsification, which is the simplest possible application problem? The answer is obvious and the solution can be constructed geometrically [17]. As shown in figure 2, we need at least $n_\mathrm{m} = 2$ measurements and a single input single output (SISO) linear model with $n_\mathrm{k} = 2$ parameters. In the sketched example, for the exactly known values x_1=1 and x_2=3, we have measured $y_1 = [1.5, 2.5]$‡ and $y_2 = [3.5, 4.5]$. The parameters $\boldsymbol{p} = (a, b)$ of the model

$$Y_s = Y(x_s, a, b) = a \cdot x_s + b \tag{2}$$

are the slope a and the intercept b. They must lie in the parallelogram $\mathbf{R}$ shown at the right part of figure 2. For all $\boldsymbol{p} \in \mathbf{R}$ we obtain models not falsified by the data, i.e. lines intersecting both measurements. More information on the model parameters can not be obtained, due to the systematic error of the y-measurements.

If we have many measurements including a random but *no* systematic error, we can additionally obtain a frequency distribution in the parameter space. The true value $\boldsymbol{p}_\mathrm{t}$ is not necessary in the center of the region of possible model parameters $\mathbf{R}$ not falsified by the measurements. In the nonlinear case, the point $\boldsymbol{p}_0$ from a goodness of fit function optimization is not necessary a member of $\mathbf{R}$. If we have only an unsubstantiated measurement error specification, we have to take the worst case, a predominant systematic error. If we have no scientifically well founded error specification, we can still guess a valid model and look for the minimum measurement error compatible to the model, i.e. providing a non empty possible parameter region $\mathbf{R}$. For the linear case, from figure 2, we can also grasp something about the necessary properties of additional measurements required to reduce the volume of $\mathbf{R}$ and thus enhance the accuracy of the indirect measurement of $\boldsymbol{p}$: For enhancing the accuracy of the indirect measurement of $\boldsymbol{p}$, more accurate measurements at a maximum distance in x are required. The covering of the intermediate x-range by inaccurate measurements serves mainly the effort to falsify the model itself. Any instance from a multitude of inaccurate measurements may falsify a model, but it may give no contribution to the accurate indirect measurement of the model parameters. The indirect measurement itself depends on reproducible accurate measurements for well selected values of the process governing control variables.

It is misleading to specify the result of a parameter calibration itself as an interval, except for the one parameter case ($n_\mathrm{k} = 1$). The region of possible parameters $\mathbf{R}$ is not necessary simply connected or even continuous.

‡ i.e. $1.5 \le y_1 \le 2.5$.

2.1. Mathematical formulation

The next step is to obtain a mathematical formulation of our intuitive falsification reasoning. It is accomplished by defining a special membership function [18], the model falsification indicator function (MFIF) F which gives 0 (false) inside $\mathbf{R}$ and 1 (true) otherwise. In our example, we have a single input single output (SISO) model (2) and measurements $y_s \pm \Delta y_s$ at exactly known x_s. The MFIF of a *single* measurement is obvious:

$$F_s(x_s, y_s, \Delta y_s, a, b) = \begin{cases} 0 & \text{if } y_s - \Delta y_s \le Y(x_s, a, b) \le y_s + \Delta y_s \\ 1 & \text{otherwise.} \end{cases} \tag{3}$$

Due to the fact that the Δy_s are not known exactly, without loss of generality, the unit step functions in (4) can be replaced by their smooth approximations using the logistic function and using the abbreviation $Y_s = Y(x_s, a, b)$ we obtain

$$F_s(x_s, y_s, \Delta y_s, a, b) = 1 + \left\{1 + \exp\left[-100\left(\frac{y_s - Y_s}{\Delta y_s} - 1\right)\right]\right\}^{-1} - \left\{1 + \exp\left[-100\left(\frac{y_s - Y_s}{\Delta y_s} + 1\right)\right]\right\}^{-1} \tag{4}$$

Being a property of the model, any modeling accuracy ΔY_s can be taken into account by replacing the Δy_s

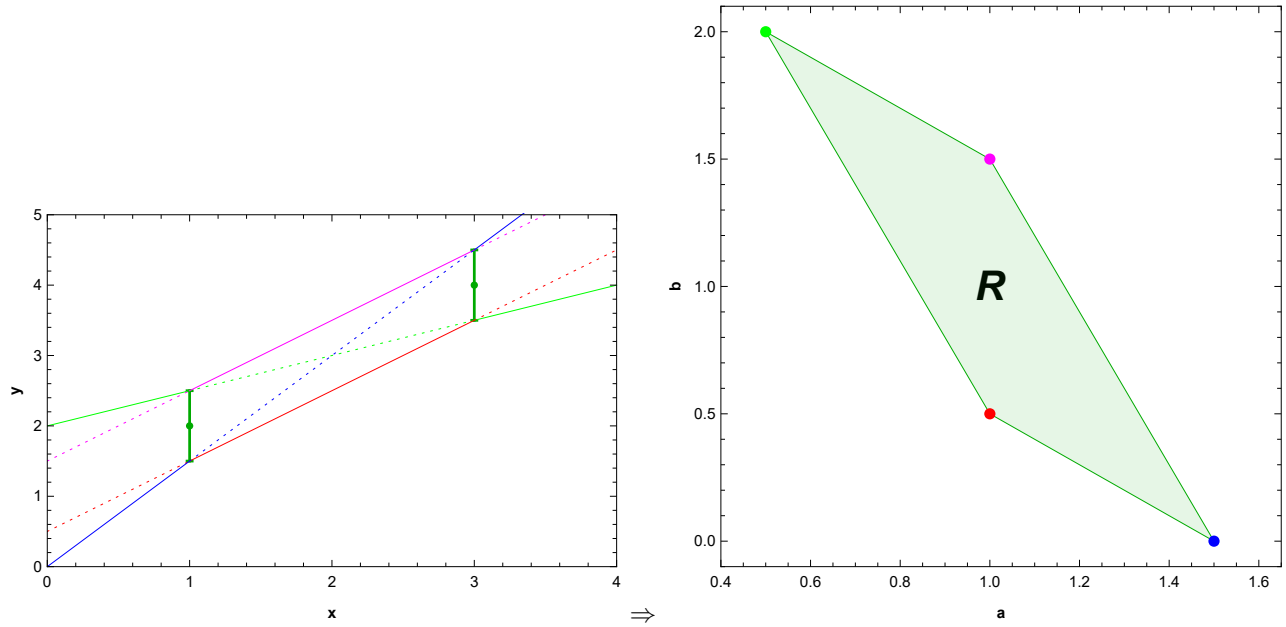

Figure 2. Linear model calibration from 2 measurements with a systematic error.

by the larger $\Delta\tilde{y}_s = \Delta y_s + \Delta Y_s$ in (4) and (5). For considering a large uncertainty of the Δy_s estimates, F_s may be modified to generate a fuzzy step (inspired by Zadeh's fuzzy logic [19]), e.g.

$$\hat{F}_s\left(x_s, y_s, \Delta y_s, a, b\right) = \tag{5}$$

$$\text{Min}\left[1, \text{Max}\left(0, \left|\frac{y_s - Y_s}{\Delta y_s}\right| - \hat{\mu}\right)\right] \tag{6}$$

or any continuously differentiable approximation of it, with the fuzziness range $\hat{\mu} \geq 1$ in multiples of Δy_s.

Assuming the trueness of the model, every individual F_s-function indicates the region in the parameter space, where the measurement s is an outlier for a specific $\boldsymbol{p}$. The model calibration has to identify the region $\mathbf{R}$ of all possible parameters, here $\boldsymbol{p} = \{a, b\} \in \mathbf{R} \subset \mathbb{R}^2$, where

$$\underset{s}{\forall}\ F_s\left(x_s, y_s, \Delta y_s, a, b\right) = 0 \tag{7}$$

For this purpose, we can define the logical MFIF $\tilde{F}$ for the complete set of the n_m measurements

$$\tilde{F}(a,b) = \text{Min}\left[1, \sum_{s=1}^{n_\text{m}} F_s\left(x_s, y_s, \Delta y_s, a, b\right)\right] \tag{8}$$

and the arithmetic mean $\overline{F}$ of the n_m measurements

$$\overline{F}(a,b) = \frac{1}{n_\text{m}} \sum_{s=1}^{n_\text{m}} F_s\left(x_s, y_s, \Delta y_s, a, b\right) \tag{9}$$

finally provides, what we may call *the* MFIF.

For visualizing the region of possible model parameters the model identification indicator function (MIIF)

$$\overline{T} = 1 - \overline{F} \tag{10}$$

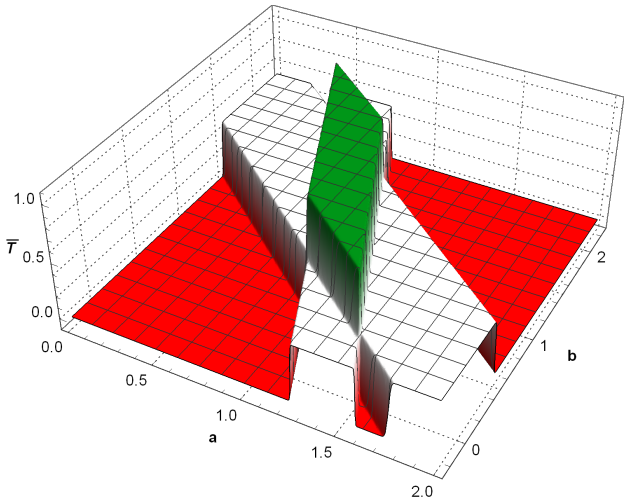

Figure 3. Model identification indicator function (MIIF) $\overline{T}$=1 − $\overline{F}$ for the line-fit problem shown in figure 2.

can be defined. Figure 3 shows $\mathbf{R}$ as the intersection of the two $\mathbf{R}_s$ from the individual measurements. Our model is unidentifiable by a single measurement, i.e. the individual $\mathbf{R}_s$ are unlimited. The flatness of the MFIF/MIIF-functions distinguishes them from the parabolic goodness of fit functions. The random part of the measurement error (precision) has a probability maximum and thus a goodness of fit function, which is science based in the case of predominating random errors, can have a minimum. The MFIF is science based in the case of predominating systematic errors and can not mark a specific parameter set, only the region of possible parameters.

$\overline{F}$ from (9) gives the fraction of outliers as a function of the model parameter vector $\boldsymbol{p} = \{a, b\}$. We can start to identify $\mathbf{R}$ by identifying the region $\mathbf{R}_\epsilon$ of all parameters $\boldsymbol{p} \in \mathbb{R}^2$ where for an $\epsilon \geq \epsilon_{\min} = 1/(2 \cdot n_\mathrm{m})$

$$\overline{F}(a, b) \leq \epsilon \tag{11}$$

If that region R_ϵ is too small to be identified, we can increase ϵ, replace F_s by $\hat{F}_s$ in (7) and (9) or we have to accept the evidence: The model Y is falsified. Additionally, for any $\epsilon \geq 3/(2 \cdot n_\mathrm{m})$, we can identify and consolidate the outliers from the set of measurements for all $\boldsymbol{p} \in \mathbf{R}_\epsilon$. Finally we get an acceptable approximation of $\mathbf{R} \approx \mathbf{R}_{\epsilon_{\min}}$. For the simplest possible case, a single input single output model with an exactly known variable x and only errors in y to be considered, we now have a mathematical method for parameter identification by falsification. In the next section, we will define and discuss the more general cases.

3. Model falsification indicator functions

Given a vector of variables

$$\boldsymbol{x} = (x_1, \ldots, x_i, \ldots, x_{n_\mathrm{i}}) \tag{12}$$

and a vector of real valued parameters

$$\boldsymbol{p} = (p_1, \ldots, p_k, \ldots, p_{n_\mathrm{k}}) \tag{13}$$

we define a Model $\boldsymbol{Y}(\boldsymbol{x}, \boldsymbol{p})$ as the mapping of the model inputs $(\boldsymbol{x}, \boldsymbol{p})$ to the outputs (MIMO=multiple inputs multiple outputs)

$$\boldsymbol{y} = \left(y_1, \ldots, y_j, \ldots, y_{n_\mathrm{j}}\right) = \boldsymbol{Y}(\boldsymbol{x}, \boldsymbol{p}). \tag{14}$$

Without loss of generality (see section 3.2), we restrict ourselves to $n_\mathrm{j} = 1$, the (MISO=multiple inputs single output)-case using $\boldsymbol{y} = Y(\boldsymbol{x}, \boldsymbol{p})$.

As discussed in section 3.3, time dependencies need no specific consideration into our formalism. Our n_m measurements include information on the accuracies (systematic errors) and precisions (random errors) which can be consolidated for each individual measurement s by the specification of the n_i intervals for the controls

$$\boldsymbol{x}_s \pm \boldsymbol{\Delta x}_s = (\ldots, [x_{i,s} - \Delta x_{i,s}, x_{i,s} + \Delta x_{i,s}], \ldots) \tag{15}$$

and the interval of the measured system response

$$[y_s - \Delta y_s, y_s + \Delta y_s]. \tag{16}$$

For any parameter vector $\boldsymbol{p}$ and each vector of measured input intervals $\boldsymbol{x}_s \pm \boldsymbol{\Delta x}_s$, it is possible to compute the interval of model outputs $Y_s \pm \Delta Y_s$

$$[Y_{s,\min}, Y_{s,\max}] = [Y_s - \Delta Y_s, Y_s + \Delta Y_s] = \tag{17}$$
$$= \text{MinMax } Y(\ldots, x_i \in [x_{i,s} - \Delta x_{i,s}, x_{i,s} + \Delta x_{i,s}], \ldots, \boldsymbol{p})$$

As most model implementations do not support interval arithmetic (as introduced by [20]), for the calculation of $Y_s \pm \Delta Y_s$ from (18) in the general (nonlinear) case, one has to use optimization algorithms or a sample of values from the hypercube defined by (15). Again, as described in section 2, any considerable error of the model itself can be added to the ΔY_s obtained from (18). Approximately, the evaluation of the model Y at the 2^{n_i} edges of the n_i dimensional hypercube defined by (15) can be used and finally for small $\boldsymbol{\Delta x}_s$ it is possible use a local linear finite difference approximation of the model obtainable from at least n_i+1 model evaluations.

For completeness, one can formulate an even more general case: If the measurements provide a bounding region in the $n_\mathrm{i} + 1$ dimensional space formed by $\boldsymbol{x}_s$ and y and the model allows to calculate a bounding region in the same space, $F_s(\boldsymbol{p})$ is defined to be 0 in the intersection of both regions and 1 otherwise. Using a (smaller) sample from the ellipsoid defined by (15) to solve (18) is insufficient.

During the model identification procedure, the experiments to be considered are fixed and we can omit the implicit arguments $\boldsymbol{x}_s$, $\boldsymbol{\Delta x}_s$, y_s and Δy_s. For a specific $\boldsymbol{p}$, we have the ability to compute Y_s and ΔY_s from (18). The definition of the model falsification indicator function (MFIF) for each individual measurement is

$$F_s(\boldsymbol{p}) = \tag{18}$$
$$\begin{cases} 0 & \text{if } y_s - \Delta y_s - \Delta Y_s \leq Y_s \leq y_s + \Delta y_s + \Delta Y_s \\ 1 & \text{otherwise.} \end{cases}$$

Analogous to (5) we obtain the smooth approximation to be

$$F_s(\boldsymbol{p}) = 1 + \tag{19}$$
$$+ \left\{1 + \exp\left[-100\left(\frac{y_s - Y_s}{\Delta y_s + \Delta Y_s} - 1\right)\right]\right\}^{-1}$$
$$- \left\{1 + \exp\left[-100\left(\frac{y_s - Y_s}{\Delta y_s + \Delta Y_s} + 1\right)\right]\right\}^{-1}$$

and the fuzzy step MFIF (6) as

$$\hat{F}_s(\boldsymbol{p}) = \text{Min}\left[1, \text{Max}\left(0, \left|\frac{y_s - Y_s}{\Delta y_s + \Delta Y_s}\right| - \hat{\mu}\right)\right] \tag{20}$$

with the fuzziness range $\hat{\mu} \geq 1$. Assuming the truth of the model, every individual F_s-function also indicates whether the measurement s for a specific $\boldsymbol{p}$ is an outlier or not. The model calibration has to identify the region $\mathbf{R}$ of all parameters $\boldsymbol{p} \in \mathbb{R}^{n_\mathrm{k}}$ where

$$\underset{s}{\forall} F_s(\boldsymbol{p}) = 0 \tag{21}$$

The logical MFIF

$$\tilde{F}(\boldsymbol{p}) = \text{Min}\left[1, \sum_{s=1}^{n_\mathrm{m}} F_s(\boldsymbol{p})\right] \tag{22}$$

has an important property when considering new experiments or when the intersection of two regions

fulfilling $\tilde{F}(\boldsymbol{p}) = 0$ for different sets of experiments is required: The MFIF's of two sets of measurements can be added analogously to (22) and the MFIF of the set union is obtained. If the MFIF for a set of measurements is computed for a predefined sample $(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_{n_\mathrm{s}})$ of the parameter space, only those points of the sample have to be stored, where the MFIF vanishes. They provide a representation of $\mathbf{R}$ for that set of measurements and using the stored sample for the next set of measurements, the intersection of the regions is obtained automatically (see also section 3.5).

Finally, analogous to (9), we define the arithmetic mean $\overline{F}$ as *the* model falsification indicator function (MFIF)

$$\overline{F}\left(\boldsymbol{p}\right) = \frac{1}{n_\mathrm{m}} \sum_{s=1}^{n_\mathrm{m}} F_s\left(\boldsymbol{p}\right) \tag{23}$$

and the model identification indicator function (MIIF)

$$\overline{T}\left(\boldsymbol{p}\right) = 1 - \overline{F}\left(\boldsymbol{p}\right). \tag{24}$$

3.1. MFIF's for exactly known controls

For the case of predominating errors in the measured y_s, ΔY_s can be neglected and Y_s can be computed exactly by a single model invocation. The MFIF specification starts with the generalization of (4)

$$F_s\left(\boldsymbol{p}\right) = \tag{25}$$
$$\begin{cases} 0 & \text{if } y_s - \Delta y_s \leq Y_s\left(\boldsymbol{x}_s, \boldsymbol{p}\right) \leq y_s + \Delta y_s \\ 1 & \text{otherwise.} \end{cases}$$

For the definitions of the other MFIF's, as $\boldsymbol{\Delta x}_s$ is neglected, one can omit ΔY_s in (19) to (20). Neglecting all $\boldsymbol{\Delta x}_s$ is the only way to avoid the (approximate) solution of the problem formulated in (18) and thus requiring only one invocation of the model per $(\boldsymbol{p}, \boldsymbol{x}_s)$.

3.2. MFIF's for the general MIMO model case

In more complex situations, we have to deal with a set of multiple (often correlated) model outputs, the MIMO-case defined in (14). Defining the MFIF requires the consideration of all y_j where measurements are available, but – as in the in-stationary case – we can consider the different y_j by the enumeration of the measurements or we define a $F_{s,j}$ for each y_j and link them together to

$$F_s\left(\boldsymbol{p}\right) = \begin{cases} 0 & \text{if } \underset{j}{\forall}\ F_{s,j}\left(\boldsymbol{p}\right) = 0 \\ 1 & \text{otherwise.} \end{cases} \tag{26}$$

For computational efficiency we have to consider that a single experiment provides multiple measurement values and a single model invocation provides all y_j per given $(\boldsymbol{p}, \boldsymbol{x}_s)$.

3.3. MFIF's for in-stationary processes

The process (and the model) can have time dependent control variables $\boldsymbol{x}(t)$ and observables $\boldsymbol{y}(t)$. Such non-stationary processes can be described by an in-stationary model $Y : \{\boldsymbol{p}, \boldsymbol{x}(t)\} \rightarrow \boldsymbol{y}(t)$. The measurements usually include some sampling in the time direction and the time can be assumed to be exactly known (synchronized clocks). It is therefore straightforward to consider the time dependencies by the enumeration of the measurements and the MFIF definitions will look the same. Now a single experiment provides even more measurement values and it is important to compute all required samples of $\boldsymbol{y}(t)$ within a single model invocation.

3.4. Model parameters

Like the goodness of fit functions, the MFIF's depend on the model parameters. The shape and volume of the feasible region $\mathbf{R}$ in the n_k dimensional model parameter space can be changed by transforming the parameters [21]. This is crucial if $\mathbf{R}$ seems to be practically boundless and mandatory for unidentifiable models ($\mathbf{R}$ becomes unlimited). In the later case, the parameter transformation has to reduce the number of parameters. In the best case, such a transformation can provide uncorrelated parameters and a new $\mathbf{R}$ closer to an ellipsoid having all major axes parallel to the coordinate axes in the parameter space. The equivalence of models depends on their identifiability and thus the accuracy of the measurements [11, 22]. The region of feasible model parameters $\mathbf{R}$ also indicates the model identifiability as introduced by Bellman and Åström [23]. For black-box or complex models, the relationships between parameter sets showing almost identical model predictions are not obvious. The model identifiability or sloppiness has to be investigated in detail [11, 24], a task which also can be performed using the MFIF.

3.5. Application of the MFIF and representation of $\mathbf{R}$

The MFIF provides a solid foundation for the identification of the region $\mathbf{R}$ of possible model parameters not falsified by a given set of measurements. The *curse of dimensionality* entails practical limitations for big n_k. We have to put as much effort in model simplification as possible (Occam's razor). For $n_\mathrm{k} \leq 3$, $\mathbf{R}$ can be visualized using standard computational tools like Mathematica. For $n_\mathrm{k} > 3$ the problem becomes increasingly difficult. Only lower dimensional projections can be visualized and the computational effort rises. In order to decrease the entry threshold, a simple algorithm for obtaining a representation of $\mathbf{R}$ is sketched here. Alternatives can be derived from the model identification

literature, e.g. SIVIA [5, 25], LSCR [26] and EDMF [12, 27].

The basic assumption is that model invocation dominates the computational effort. The approach is to represent $\mathbf{R}$ in the n_k-dimensional parameter space by a representative set of n_r points $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_l, \ldots, \boldsymbol{r}_{n_\mathrm{r}})$. First of all, we need an initial $\boldsymbol{p}_0$ where the model is not falsified by the measurements, i.e. $\overline{F}(\boldsymbol{p}_0) \leq \epsilon_\mathrm{min} = 1/(2{\cdot}n_\mathrm{m})$ or $\overline{F}(\boldsymbol{p}_0) \leq \epsilon$, if we want to allow for a certain number of outliers in the set of measurements

$$\epsilon = (2 \cdot n_\mathrm{outliers} + 1) / (2 \cdot n_\mathrm{m}) . \tag{27}$$

$\boldsymbol{p}_0$ can be obtained by any crude solution of

$$\boldsymbol{p}_0 = \arg\min \overline{F}(\boldsymbol{p}) \tag{28}$$

or from (1), if $\overline{F}$ makes numerical problems for a specific optimization algorithm. In the next step we need a vector of n_u candidate points $(\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{n_\mathrm{u}})$ in the n_k-dimensional space, e.g. located in a (scaled) sphere of radius d around $\boldsymbol{p}_0$

$$\underset{\boldsymbol{u}}{\forall} \; ||\boldsymbol{u} - \boldsymbol{p}_0|| \leq d \tag{29}$$

From numerical integration in high dimensional spaces, proper sampling methods for the generation of n_v candidate points $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n_\mathrm{v}}\}$ in the n_k-dimensional space are known [28] and available in standard computational environments like Mathematica. Initially and only once for a specific dimension n_k, one has to sample a sufficiently large n_v for getting $n_\mathrm{u} \ll n_\mathrm{v}$ points fulfilling (29). Finally we get $n_\mathrm{r} \ll n_\mathrm{u}$ points representing $\mathbf{R}$, i.e.

$$\underset{\boldsymbol{r}}{\forall} \; \overline{F}(\boldsymbol{r}) \leq \epsilon. \tag{30}$$

For a further reduction of the computational effort we have to use adaptive methods, see e.g. [5, 9]. Finally, having a representation of $\mathbf{R}$ by a finite set of points in the parameter space, we can use this set for all the purposes of model identification:

- visualization of the possible parameter space and its specific drawbacks, e.g. model sloppiness, correlated parameters and practical unidentifiabilities.
- identification of additional experiments with maximum impact on a further reduction of $\mathbf{R}$.
- reduction of the possible parameter space by the intersection of $\mathbf{R}$'s from different experiments.
- $\mathbf{R}$ as a representation of all possible parameter sets currently not falsified by the available experimental data can be used for process optimization purposes.

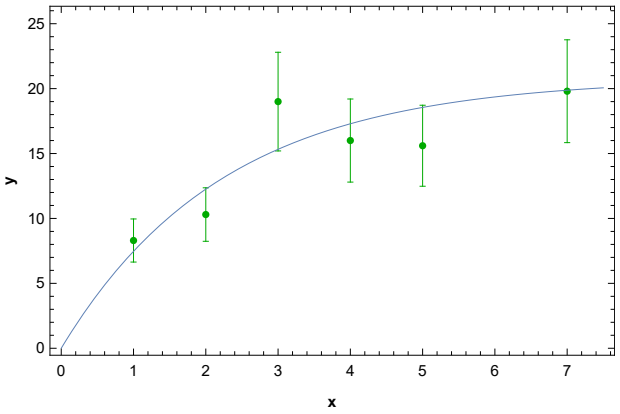

Figure 4. The BOD measurements of table 1 and the model of equation (31) for a parameter set not falsified by the measurements assuming an accuracy of 20%.

x	1.	2.	3.	4.	5.	7.
y	8.3	10.3	19.	16.	15.6	19.8

Table 1. BOD data of the nonlinear calibration example (after [29, p.72]).

4. Nonlinear example

In section 2, the model falsification indicator function was obtained for a linear regression problem. In this section, the applicability to a non-linear problem will be demonstrated for an easily reproducible example. It is taken from [29, p.72] and often used for testing nonlinear regression methods, see e.g. [30]. It is using the data from a biochemical oxygen demand (BOD) experiment carried out by Marske in the year 1967. By omitting all units and as shown in table 1, we have $n_\mathrm{m} = 6$ (x, y)-tuples and we have to treat the unknown measurement accuracy Δy_s as the parameter of our investigation.

For the 2-parameter model

$$Y_s = Y(x_s, a, b) = a \cdot \left(1 - \mathrm{e}^{-b \cdot x_s}\right) \tag{31}$$

in figure 4, the solution of the inverse problem is shown. Taking the y-values of table 1, the task was to determine the smallest systematic errors compatible with the assumption of (31) being the correct model not falsified by the data. The solution is found to be approximately $\boldsymbol{p} = (a, b) = (20.8, 0.445)$ for an error above 19.5%. As discussed in section 3.1 we neglect the Δx-error ($\Rightarrow \Delta Y_s = 0$) and use F_s as defined in (20) to calculate the arithmetic mean model falsification indicator function $\overline{F}(\boldsymbol{p})$ from (23). As we have $n_\mathrm{m} = 6$ measurements and we are only interested in all parameters having not more than one outlier ($n_\mathrm{outliers} \leq 1$) we have to look for $\overline{T}(\boldsymbol{p}) = 1 - \overline{F}(\boldsymbol{p}) \in [\frac{3}{4}, 1]$ as shown in figure 5 for a measurement accuracy of 25%. The high plateau (at $\overline{T}$=1) in figure 5 denotes the region of possible model parameters not

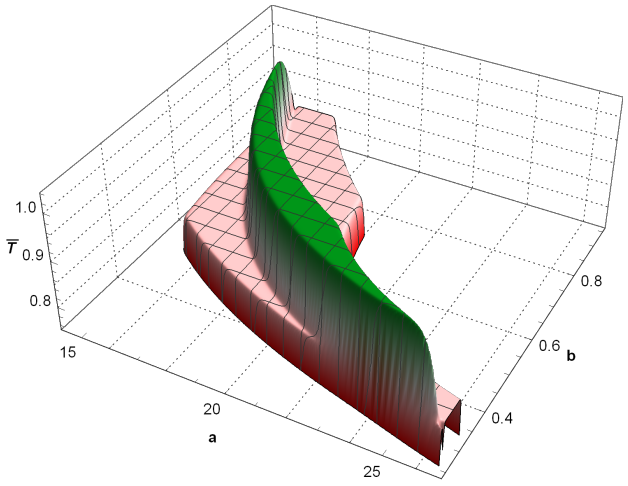

Figure 5. Model identification indicator function $\overline{T}$ for the BOD measurements of table 1, the model defined by (31) and a measurement accuracy of 25%.

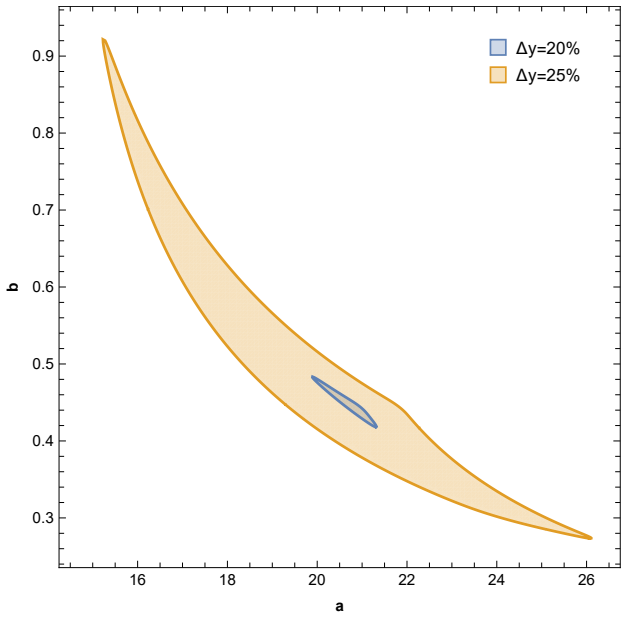

Figure 6. Regions of possible model parameters for a measurement accuracy of 20% and 25%.

falsified by the data for a measurement error of 25%. The step below (at $\overline{T}$=1-$\frac{1}{6}$) denotes all parameters leading to exactly one outlier in the data. The regions of possible model parameters can be identified by $\overline{F} \leq 1/(2 \cdot n_{\mathrm{m}}) = \frac{1}{12}$ and are plotted in figure 6 for measurement accuracies of 20% and 25%. Reducing the parameter of our consideration, the *measurement accuracy* Δy, below 20% gives an ellipsoid vanishing at $\boldsymbol{p} = (a, b) = (20.8, 0.445)$ for an error below 19.5%. The results demonstrate the applicability of the MFIF for model calibration in the nonlinear case. For the specific BOD example, it is obvious that more accurate measurements are required. The results are at least as feasible as those from other methods [29, 30] and demonstrate the low entry threshold of the MFIF-method.

5. Summary and conclusion

For multiple parameter models, measurement errors determine the size of the region of possible model parameters. It is misleading to specify the model parameters as intervals because the outcoming hyperrectangle includes much more than the region of possible model parameters. For this region there are two limiting cases:

(i) In the case of predominating *random* errors the method of least squares (LSM) delivers the most probable parameter set and its statistical enhancement the corresponding PDF in the parameter space [1, 8].

(ii) In the case of predominating *systematic* errors the bounding approaches, as discussed in this paper, deliver the region of possible parameters.

In practical applications, LSM can deliver an initial guess and the MFIF will allow to compute a representative sample of the region of possible parameters. Compared to alternative approaches, e.g. based on model linearisation and the characterization of the parameter region by large systems of interval inequalities, the MFIF approach is simple and seems to have a minimum computational complexity. The MFIF replaces the goodness of fit functions for the case of predominating systematic errors.

Being accepted, it will force experimentalists to specify their measurement errors because these are as important an ingredient as the measurement values itself. Considering random errors will raise up the next question: All $\boldsymbol{p} \in \mathbf{R}$ are equal or some are more equal? Here, all the advanced statistical methods can find their specific applications.

The model falsification indicator function (MFIF) does not require any statistical assumptions. All measurements are quantified by intervals. The method is safe and obviously well founded. For new applications, a significantly lower manpower effort is anticipated.

The simplest possible application (linear model fit) and a simple but instructive nonlinear example were used to exemplify all definitions.

MFIF based methods will transform the perception of the falsification method: It is constructive not destructive! The positive quantitative application of model falsification was established and exemplified here.

For the individual scientist, it may be difficult to accept: A single measurement, excluded to be

an outlier by independent reproduction, falsifies a model and enforces the development of a better one. For the feasibility of science, this is a prerequisite. Additionally, according to our limited ability to grasp the laws of nature or compute their consequences, we can quantify the accuracy of our models.

The MFIF simplifies the determination of the region of all possible parameters for all kinds of models. By considering the errors in the inputs and a representative sample of the parameters, *reliable modeling* requires just a bit more computing capacity.

References

[1] M Grabe. *Measurement uncertainties in science and technology*. Springer, 2 edition, 2014. DOI:10.1007/978-3-319-04888-8.

[2] K Popper. *Logik der Forschung*. Number 9 in Schriften zur wissenschaftlichen Weltauffassung. Springer, Wien, 1935.

[3] A Tarantola. Popper, bayes and the inverse problem. *Nature Physics*, 2:492–494, Aug. 2006. DOI:10.1038/nphys375.

[4] J V Beck and K J Arnold. *Parameter estimation in engineering and science*. Wiley, New York, 1977. (Google).

[5] E Walter and L Pronzato. *Identification of parametric models from experimental data*. Springer, London, 1997. (Google).

[6] L Ljung. *System Identification: Theory for the user*. Prentice Hall, 1999. (Google).

[7] S M Stigler. Gauss and the invention of least squares. *The Annals of Statistics*, 9(3):465–474, 1981. www.jstor.org/stable/2240811.

[8] B C Csaji, M Campi, and E Weyer. Non-asymptotic confidence regions for the least-squares estimate. *IFAC Proceedings Volumes*, 45(16):227–232, 2012. DOI:10.3182/20120711-3-BE-2027.00417.

[9] O Klepper and E M T Hendrix. A comparison of algorithms for global characterization of confidence regions for nonlinear models. *Environmental Toxicology and Chemistry*, 13(12):1887–1899, 1994. DOI:10.1002/etc.5620131204.

[10] M Kieffer and E Walter. Guaranteed characterization of exact non-asymptotic confidence regions as defined by lscr and sps. *Automatica*, 50(2):507–512, Feb 2014. DOI:10.1016/j.automatica.2013.11.010.

[11] D V Raman, J Anderson, and A Papachristodoulou. Delineating parameter unidentifiabilities in complex models. *Physical Review E*, 95(3):32314, 2017. DOI:10.1103/PhysRevE.95.032314.

[12] J A Goulet and I F C Smith. Structural identification with systematic errors and unknown uncertainty dependencies. *Computers and Structures*, 128:251–258, 2013. DOI:10.1016/j.compstruc.2013.07.009.

[13] R E Moore. *Methods and Applications of Interval Analysis*. siam, Philadelphia, 1979. DOI:10.1137/1.9781611970906 .

[14] E Walter and H Piet-Lahanier. Estimation of parameter bounds from bounded-error data: a survey. *Mathematics and Computers in Simulation*, 32(5-6):449–468, 1990. DOI:10.1016/0378-4754(90)90002-Z.

[15] M Kieffer and E Walter. Guaranteed estimation of the parameters of nonlinear continuous-time models: Contributions of interval analysis. *International Journal of Adaptive Control and Signal Processing*, 25(3):191–207, 2011. DOI:10.1002/acs.1194.

[16] M Henrion and B Fischhoff. Assessing uncertainty in physical constants. *American Journal of Physics*, 54(9):791–798, 1986. DOI:10.1119/1.14447.

[17] O Artbauer. Application of interval, statistical, and fuzzy methods to the evaluation of measurements. *Metrologia*, 25(2):81–86, 1988. DOI:10.1088/0026-1394/25/2/004.

[18] H Bandemer. *Mathematics of Uncertainty*, volume 189 of *Studies in Fuzziness and Soft Computing*. Springer, Berlin, 2006. DOI:10.1007/3-540-31228-5.

[19] L A Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965. DOI:10.1016/S0019-9958(65)90241-X.

[20] R E Moore, R B Kearfott, and M J Cloud. *Introduction to interval analysis*. siam, Philadelphia, 2009. DOI:10.1137/1.9780898717716.

[21] G J S Ross. *Nonlinear Estimation*. Springer Series in Statistics. Springer, New York, 1990. DOI:10.1007/978-1-4612-3412-8.

[22] I Stelzl. Changing a causal hypothesis without changing the fit: some rules for generating equivalent path models. *Multivariate Behavioral Research*, 21(3):309–331, 1986. DOI:10.1207/s15327906mbr2103_3.

[23] R Bellman and K J Åström. On structural identifiability. *Mathematical Biosciences*, 7(3-4):329–339, 1970. DOI:10.1016/0025-5564(70)90132-X.

[24] E Walter. *Identifiability of Parametric Models*. Pergamon Press, Oxford, 1987. DOI:10.1016/C2013-0-03836-4.

[25] L Jaulin and E Walter. Set inversion via interval analysis for nonlinear bounded-error estimation. *Automatica*, 29(4):1053–1064, 1993. DOI:10.1016/0005-1098(93)90106-4.

[26] M Campi and E Weyer. Guaranteed non-asymptotic confidence regions in system identification. *Automatica*, 41(10):1751–1764, 2005. DOI:10.1016/j.automatica.2005.05.005.

[27] R Pasquier and I F C Smith. Robust system identification and model predictions in the presence of systematic uncertainty. *Advanced Engineering Informatics*, 29(4):1096–1109, 2015. DOI:10.1016/j.aei.2015.07.007.

[28] I M Sobol. On the systematic search in a hypercube. *SIAM Journal on Numerical Analysis*, 16(5):790–793, 1979. DOI:10.1137/0716058.

[29] D M Bates and D G Watts. *Nonlinear regression analysis and its applications*. Wiley, New York, 2007. DOI:10.1002/9780470316757.

[30] M Schwaab. Nonlinear parameter estimation through particle swarm optimization. *Chemical Engineering Science*, 63(6):1542–1552, 2008. DOI:10.1016/j.ces.2007.11.024.